Smoothing for non-smooth optimization, lecture 1

Instructor:

Javier Peña, Carnegie Mellon University

Goals:

- First-order (gradient-based) algorithms for huge convex optimization problems
- Two applications:
  - Game theory: $\max_{y \in Q_2} \min_{x \in Q_1} \langle Ax, y \rangle$
  - Sparse principal components analysis:

$$\max_{X \in Q_1} \min_{Y \in Q_2} \{\langle C, X \rangle - \langle X, Y \rangle\}$$

About algorithms for convex optimization

- Interior-point methods: based on Newton's method. $O(\log(1/\epsilon))$ expensive iterations to find ϵ-approximate solution
- Subgradient methods (non-smooth case) $O(1/\epsilon^2)$ cheap iterations to find ϵ-approximate solution
- Efficient gradient-based methods (smooth case) $O(1/\sqrt{\epsilon})$ cheap iterations to find ϵ-approximate solution
- Smoothing techniques (non-smooth case) $O(1/\epsilon)$ cheap iterations to find ϵ-approximate solution

Course outline:

- Introduction & motivation
- Basics of convex analysis
- Gradient and subgradient schemes
- Smoothing techniques
- Applications

Brief, incomplete, and biased history of convex optimization

- 19th century: optimization in physics, Gauss's least-squares method
- 1900-1970: math developments, convex optimization
- 1940s: simplex method for linear programming
- 1970s: ellipsoid method
- 1980s: interior-point methods
- 1990s: algorithms for semidefinite programming, and symmetric cone programming, new applications in stats, machine learning, combinatorial optimization, control, circuit design, etc.
- 2000s: methods with low computational cost for huge problems

Convex optimization

$$\begin{array}{rl} \text{min} & f(x) \\ \text{s.t.} & x \in D \end{array}$$

where

- f: convex function
- D: convex set

Examples of convex optimization

- Least squares

$$\min_x \|Ax - b\|^2$$

- Linear programming

$$\begin{array}{rl} \text{min} & c^\mathsf{T} x \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{array}$$

- Semidefinite programming

$$\begin{array}{rl} \text{min} & C \bullet X \\ \text{s.t.} & \mathcal{A}X = b \\ & X \succeq 0 \end{array}$$

Here the variable X is an $n \times n$ symmetric matrix and

$$X \succeq 0 \Leftrightarrow u^\mathsf{T} X u \geq 0 \; \forall u \in \mathbb{R}^n \Leftrightarrow \lambda(X) \geq 0.$$

Convention: $\lambda(X) = (\lambda_1(X), \ldots, \lambda_n(X))$ eigenvalues of X with

$$\lambda_1(X) \geq \cdots \geq \lambda_n(X).$$

Today:

- Convex sets
- Convex functions
- Smooth convex functions
- Non-smooth convex functions
- Fenchel duality

Convex sets

Throughout this course

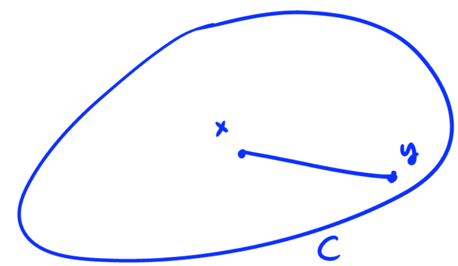
E: finite-dimensional Euclidean space with inner product $\langle \cdot, \cdot \rangle$.

$\| \cdot \|$: norm induced by $\langle \cdot, \cdot \rangle$.

Two special cases:

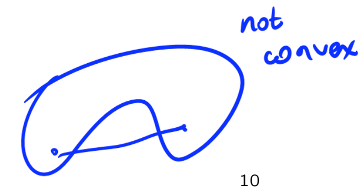
- $\mathbb{R}^n$, with inner product $\langle x, s \rangle = x^{\mathsf{T}} s$
- $\mathbf{S}^n$: space of $n \times n$ real symmetric matrices with inner product $\langle X, S \rangle = \text{trace}(XS) =: X \bullet S$.

Recall: $\text{trace}(M) = \sum_i \lambda_i(M) = \sum_i M_{ii}$.



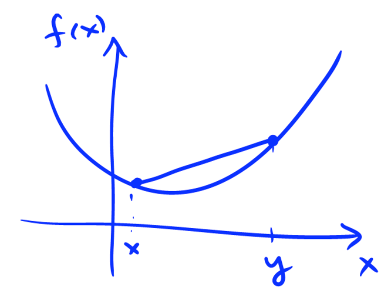
Definition 1

- $C \subseteq E$ is *convex* if $\alpha x + (1-\alpha) y \in C$ whenever $x, y \in C$ and $\alpha \in [0, 1]$.
- $K \subseteq E$ is a cone if $\lambda x \in K$ whenever $x \in K$ and $\lambda \geq 0$.



Example 2

- Polyhedron: $\{x \in \mathbb{R}^n : Ax \leq b\}$ for some given $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$.
- Non-negative orthant: $\mathbb{R}^n_+ := \{x \in \mathbb{R}^n : x \geq 0\}$
- Polytope: $\{My : y \geq 0, \sum_{j=1}^n y_j = 1\}$ for some given $M \in \mathbb{R}^{n \times m}$.
- Simplex: $\Delta_n := \{x \in \mathbb{R}^n : x \geq 0, \sum_{i=1}^n x_i = 1\}$
- Ball: $\left\{x \in E : \langle x, x \rangle \leq r^2\right\}$ for some given $r > 0$.
- Positive semidefinite cone: $\mathbf{S}^n_+ := \{X \in \mathbf{S}^n : \lambda(X) \geq 0\}$
- Spectraplex: $\left\{X \in \mathbf{S}^n_+ : \sum_{i=1}^n \lambda_i(X) = 1\right\}$.



Convex functions

Definition 3 Let $C \subseteq E$ be a convex set and $f : C \to \mathbb{R}$. The function f is convex on C if for all $x, y \in C$ and $\alpha \in [0, 1]$

$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y).$$

Alternative definition of convex functions

- Consider extended-valued functions $f : E \to \mathbb{R} \cup \{+\infty\}$.
- Can always do this by declaring $f(x) = +\infty$ for $x \notin \text{dom}(f)$
- f is convex if and only its *epigraph*

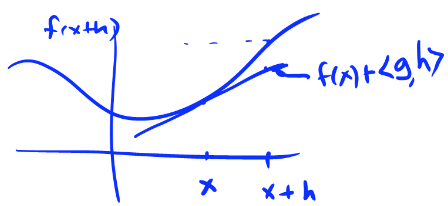
$$\text{epi}(f) := \{(x, r) \in E \times \mathbb{R} : r \geq f(x)\},$$

  is a convex set.

- Define $\text{dom}(f) := \{x \in E : f(x) < \infty\}$ (projection of $\text{epi}(f)$)

Example 4 Some convex functions

- $f : E \to \mathbb{R}$ defined by $x \mapsto \langle x, x \rangle$ $= \|x\|^2$
- $f : E \to \mathbb{R}$ defined by $x \mapsto \langle a, x \rangle$ for some $a \in E$
- $f : E \to \mathbb{R} \cup \{+\infty\}$ defined by $x \mapsto \sup_{i \in I} g_i(x)$ where each $g_i : E \to \mathbb{R} \cup \{+\infty\}$, $i \in I$ is convex
- $f : \mathbb{R} \to \mathbb{R}$ defined by $x \mapsto e^x$
- $f : \mathbb{R}_+ \to \mathbb{R}$ defined by $x \mapsto x \log x$
- $f : \mathbb{S}^n \to \mathbb{R}$ defined by $X \mapsto \lambda_1(X) + \cdots + \lambda_k(X)$, where $k \leq n$



Smooth convex functions

Definition 5 $f : E \to \mathbb{R} \cup \{+\infty\}$ is *differentiable* at $x \in \text{dom}(f)$ if there exists $g \in E$ such that

$$\lim_{h \to 0} \frac{f(x+h) - f(x) - \langle g, h \rangle}{\|h\|} = 0,$$

In other words, if

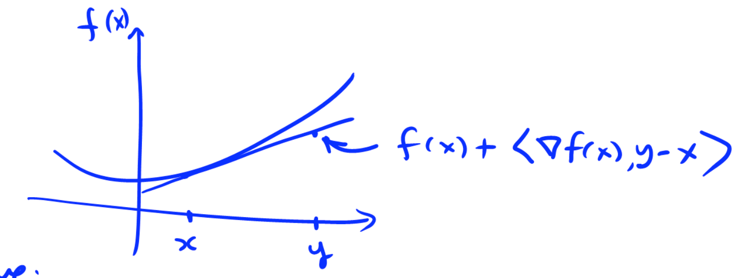
$$f(x+h) = f(x) + \langle g, \overset{h}{\cancel{x}} \rangle + o(\|h\|).$$

When such g exists, it is unique. In this case, g is the *gradient* of f at x, and is denoted $\nabla f(x)$.

Proposition 6 *Assume f is differentiable on an open set $\Omega \subseteq E$ and $C \subseteq \Omega$ is convex. Then the following are equivalent*

- *f is convex on C*
- $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$ for all $x, y \in C$
- $\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0$ for all $x, y \in C$

Corollary 7 *Assume $f : E \to \mathbb{R}$ is differentiable everywhere and convex. Then $\bar{x} = \text{argmin}_x f(x)$ if and only if $\nabla f(\bar{x}) = 0$.*



To prove:

$$f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle$$

Pf. Consider $\varphi(t) := f(x+t(y-x))$

$$\varphi'(0) = \langle \nabla f(x), y-x \rangle$$

$$\varphi(t) = f(x+t(y-x)) = f(ty+(1-t)x) \leq tf(y)+(1-t)f(x$$

$$\leq f(x)+t(f(y)-f(x)) = \varphi(0)+t(f(y)-f(x))$$

$$[\varphi(t)-\varphi(0)]/t \leq f(y)-f(x)$$

$\nabla f : \Omega \subseteq E \to E$

∇f Lipschitz if $\|\nabla f(x)-\nabla f(y)\| \leq L\|x-y\|$

Lipschitz continuity of the gradient

Theorem 8 *Assume f is differentiable on an open set $\Omega \subseteq E$ and $C \subseteq \Omega$ is convex. Then the following are equivalent:*

- *f is convex and ∇f is Lipschitz with constant L on C*
- $0 \leq f(y) - f(x) - \langle \nabla f(x), y-x \rangle \leq \frac{L}{2}\|y-x\|^2$ *for all $x, y \in C$*
- $\frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2 \leq f(y) - f(x) - \langle \nabla f(x), y-x \rangle$ *for all $x, y \in C$*
- $\frac{1}{L}\|\nabla f(y) - \nabla f(x)\|^2 \leq \langle \nabla f(y) - \nabla f(x), y-x \rangle$ *for all $x, y \in C$*

$$\frac{1}{2L}\|\nabla f(x)-\nabla f(y)\|^2 \leq f(x)-f(y)-\langle \nabla f(y), x-y \rangle$$

Strong convexity

Definition 9 Let $C \subseteq E$ be a convex set and $f : C \to \mathbb{R}$. The function f is strongly convex on C if there exists $\rho > 0$ such that for all $x, y \in C$ and $\alpha \in [0,1]$

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) - \frac{1}{2}\rho\alpha(1-\alpha)\|x-y\|^2.$$

In this case say that f is strongly convex with modulus ρ.

Proposition 10 *Let $C \subseteq E$ be a convex set and $f : C \to \mathbb{R}$. Then f is strongly convex with modulus $\rho > 0$ if and only if $f - \frac{1}{2}\rho\|\cdot\|^2$ is convex on C.*

Theorem 11 *Assume f is differentiable on an open set $\Omega \subseteq E$ and $C \subseteq \Omega$ is convex. Then the following are equivalent*

- *f is strongly convex on C with modulus ρ*
- $f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle + \frac{1}{2}\rho\|x-y\|^2$ *for all $x, y \in C$*
- $\langle \nabla f(y) - \nabla f(x), y-x \rangle \geq \rho\|x-y\|^2$ *for all $x, y \in C$*

f strongly convex w/ modulus $\rho \iff$

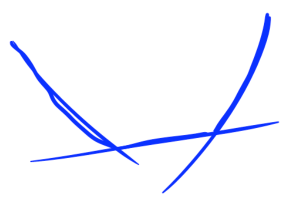
$\longrightarrow f - \frac{1}{2}\rho\|\cdot\|^2$ convex

$$f(\alpha x+(1-\alpha)y) - \tfrac{1}{2}\rho\|\alpha x+(1-\alpha)y\|^2 \le \alpha\left[f(x)-\tfrac{1}{2}\rho\|x\|^2\right] + (1-\alpha)\left[f(y)-\tfrac{1}{2}\rho\|y\|^2\right]$$

$$f(\alpha x+(1-\alpha)y) \le \alpha f(x)+(1-\alpha)f(y) - \tfrac{1}{2}\rho\underbrace{\left[\alpha\|x\|^2+(1-\alpha)\|y\|^2-\|\alpha x+(1-\alpha)y\|^2\right]}_{\alpha(1-\alpha)\|y-x\|^2}$$

Example 12 Some strongly convex functions

- $f : E \to \mathbb{R}$ defined by $x \mapsto \langle x, x\rangle$
- $f : \Delta_n \to \mathbb{R}$ defined by $x \mapsto \sum_{i=1}^{n} x_i \log x_i$
- $f : \{X \in \mathbb{S}^n : X \succeq 0,\ I \bullet X = 1\} \to \mathbb{R}$ defined by $X \mapsto \sum_{i=1}^{n} \lambda_i(X) \log \lambda_i(X)$

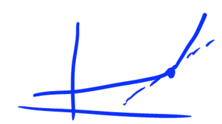


Non-smooth convex functions

Non-smoothness is pervasive in optimization.

Example 13

- $f(x) = \sup\{g_i(x) : i \in I\}$ *where* $g_i,\ i \in I$ *are convex.*
- $f(x) = \inf\{\phi(u) : c_i(u) \le x_i, i = 1, \ldots, n\}$ *where* $\phi, c_i, \ldots, c_n$ *are convex.*



Proposition 14 *Assume* $f : E \to \mathbb{R} \cup \{+\infty\}$ *convex and* $x \in \mathrm{int}(\mathrm{dom}(D))$. *Then there exists* $g \in E$ *such that*

$$f(y) - f(x) \ge \langle g, y - x\rangle \quad \text{for all} \quad y \in E. \tag{1}$$

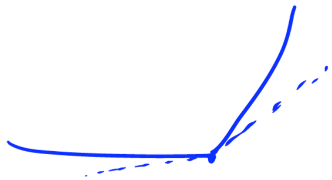
$f(y)-f(x) \ge \langle \nabla f(x), y-x\rangle$

Generalization of the gradient:

Definition 15 Assume $f : E \to \mathbb{R} \cup \{+\infty\}$ convex and $x \in \mathrm{dom}(D)$. The *subdifferential* of f at x is

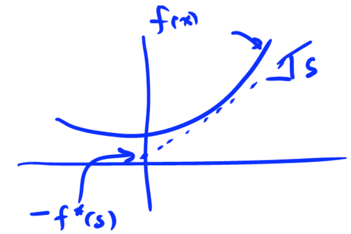
$$\partial f(x) := \{g \in E : (1) \text{ holds}\}$$

Corollary 16 *Assume* $f : E \to \mathbb{R} \cup \{+\infty\}$ *is convex. Then* $\bar{x} \in \mathrm{argmin}_x f(x)$ *if and only if* $0 \in \partial f(\bar{x})$.



Notice:
If f is convex and differentiable at x then $\partial f(x) = \{\nabla f(x)\}$.

Proposition 17 *Assume $f : E \to \mathbb{R} \cup \{+\infty\}$ convex Then f is differentiable at x if and only if $\partial f(x)$ is a singleton.*



Fenchel conjugate

Definition 18 Assume $f : E \to \mathbb{R} \cup \{+\infty\}$, with $\text{dom}(f) \neq \emptyset$. The *conjugate* of f is $f^* : E \to \mathbb{R} \cup \{+\infty\}$ defined by

$$s \mapsto f^*(s) := \sup\{\langle s, x\rangle - f(x) : x \in \text{dom}(f)\}.$$

Notice:

- $f^*(s) + f(x) \geq \langle s, x\rangle$ for all $x, s \in E$
- $f^*(s) + f(x) = \langle s, x\rangle$ if and only if $s \in \partial f(x)$

Example 19

$f(x)$	$f^*(s)$
e^x	$s \log s - s$
$\frac{\lvert x\rvert^p}{p},\ p > 1$	$\frac{\lvert s\rvert^q}{q},\ \frac{1}{q} = 1 - \frac{1}{p}$
$\sqrt{1 + x^2}$	$-\sqrt{1 - s^2}$
$-\log x$	$-1 - \log(-s)$

Definition 20 $f : E \to \mathbb{R} \cup \{+\infty\}$ is closed if $\text{epi}(f)$ is closed.

Proposition 21 *Assume $f : E \to \mathbb{R} \cup \{+\infty\}$.*

- *f^* is closed and convex*
- *$f^{**} \leq f$*
- *$f^{**} = f$ if and only if f is closed and convex.*

$f^*(s) = \sup_x \{\langle s, x\rangle - f(x)\}$

Fenchel duality

Assume $f : E \to \mathbb{R} \cup \{+\infty\}$, $g : Y \to \mathbb{R} \cup \{+\infty\}$ and $A \in L(E, Y)$.

Consider the pair

$$p := \inf_{x \in E} \{f(x) + g(Ax)\} \qquad \text{(P)}$$

and

$$d := \sup_{y \in Y} \{-f^*(A^*y) - g^*(-y)\} \qquad \text{(D)}$$

Theorem 22

- *(Weak duality):* $p \geq d$
- *(Strong duality): If* $0 \in \text{relint}(\text{dom}(g) - A\text{dom}(f))$ *then* $p = d$ *and* d *is attained.*

$$f(x) + g(Ax) + f^*(A^*y) + g^*(-y) \geq \langle x, A^*y \rangle + \langle Ax, -y \rangle = 0$$

Special case

Consider

$$\min_{x \in Q_1} \max_{y \in Q_2} \left\{ \hat{f}(x) - \hat{\phi}(y) + \langle Ax, y \rangle \right\}. \qquad (2)$$

where

- E_1, E_2 are finite dimensional Euclidean spaces, $A \in L(E_1, E_2)$
- $Q_i \subseteq E_i$ are simple compact convex sets
- $\hat{f}$ and $\hat{\phi}$ are convex and differentiable everywhere

Can write (2) as

$$\min_{x \in E} \{f(x) + \phi^*(Ax)\}$$

for

- $f(x) = \hat{f}(x)$, $\text{dom}(f) := Q_1$
- $\phi(y) = \hat{\phi}(y)$, $\text{dom}(\phi) := Q_2$

Fenchel dual:

$$\max_{y \in E_2} \{-f^*(A^*y) - \phi^{**}(-y)\}$$

which can be written as

$$\max_{y \in Q_2} \min_{x \in Q_1} \left\{ \hat{f}(x) - \hat{\phi}(y) + \langle Ax, y \rangle \right\}.$$

Fenchel duality yields

$$\min_{x \in Q_1} \max_{y \in Q_2} \left\{ \hat{f}(x) - \hat{\phi}(y) + \langle Ax, y \rangle \right\} = \max_{y \in Q_2} \min_{x \in Q_1} \left\{ \hat{f}(x) - \hat{\phi}(y) + \langle Ax, y \rangle \right\}.$$

References for today's material

- J. Borwein and A. Lewis, "Convex Analysis and Nonlinear Optimization," Springer Verlag, 2005.
- S. Boyd and L. Vanderberghe, "Convex Optimization," Cambridge Academic Press, 2004. Available from `http://www.stanford.edu/~boyd/cvxbook/`
- Y. Nesterov, "Introductory Lectures on Convex Optimization," Kluwer Academic Publishers, 2004.
- A. Ben-Tal and A. Nemirovski, "Lectures on Modern Convex Optimization," MPS-SIAM Series on Optimization, 2001. Related material available from `http://www2.isye.gatech.edu/~nemirovs/`